MQ Data Science

Report to MRC

*on building a **National Infrastructure for Mental Health Data Science** in the UK*

9th September 2019
St Paul's and St George's Church, Edinburgh

Knowledge Exchange Team
Mental Health Data Science Scotland
University of Edinburgh

Beange, I., Iveson, M., Mai, H., McIntosh, A., Xenaki, C., Zhang, N.
With Smith, D.(University of Glasgow) & Wölbert, E (MQ)

MRC Medical Research Council

THE UNIVERSITY of EDINBURGH

MQ Transforming mental health

MHDSS Mental Health Data Science Scotland

Table of contents

Introduction

At the MQ meeting on 9th September 2019 (Edinburgh), we hosted a discussion session, where we asked participants how they would envision a National Infrastructure for Mental Health Data Science (referred to as “platform” below). Participants were advised that their collected input would be fed back to the MRC to serve as a starting point for their full consultation on this topic.

Key findings

- Researchers are particularly interested in using such a platform themselves and recognise its utility to a large number of other organisations.

- The minimum useful requirement for such a platform is a searchable directory of available data with a note about access procedures. Whereas an ideal platform might contain the actual data. Our participants demonstrated a wide interest in the inclusion of different data types and research areas within such a platform.

- Participants were enthusiastic about the possibilities such a platform might enable, mentioning expansion of research topics and conducting more rigorous research. They felt the platform might enable this by maximizing data quality/usage and facilitating data access.

- Participants recognised the positive impact such a platform could have on people affected by mental illness. Patterns of responses arose around patient engagement, policy shaping, open science, and improved research quality. All these elements could help fight the stigma of mental illness in society.

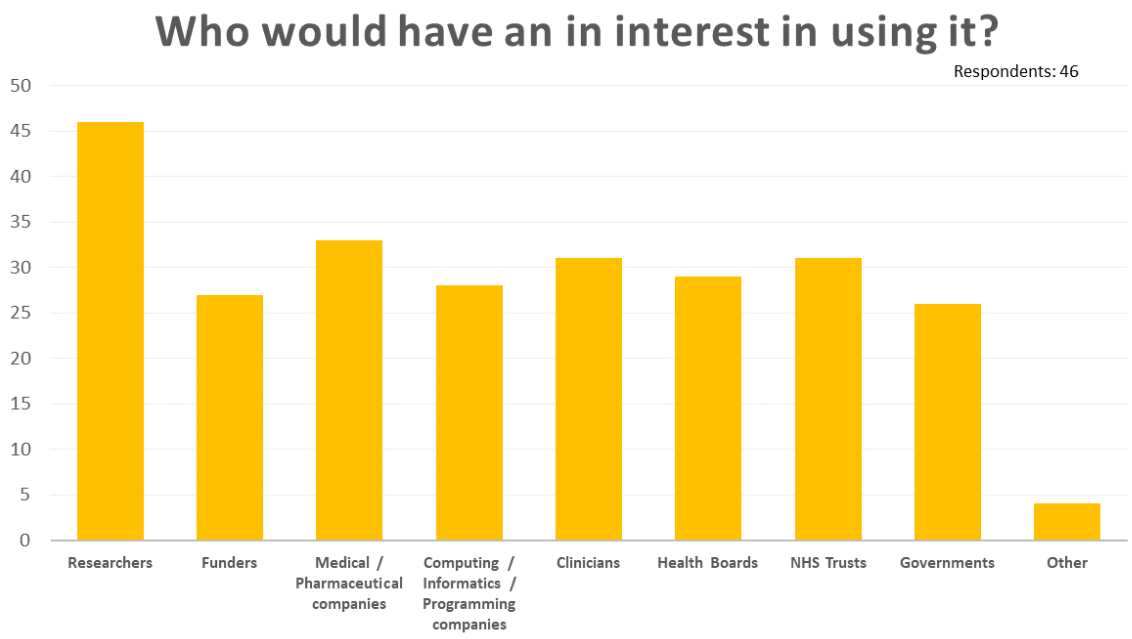
- With regard to the MRC consultation, our participants recognised the need to consult with a range of organisations, including data donors, ethical boards, privacy panels and relevant charities - with an observed less enthusiastic support for political involvement (which may reflect current political circumstances).

- Our participants raised a number of concerns, which the MRC may wish to address during their consultation phase. These included long-term sustainability, governance, and ethical considerations. On a practical note, questions were raised around prioritization, practicalities on data management, platform operation and relevant training and capacity building. Lastly, the participants illustrated the need to raise public awareness of and involvement in this platform in order to ensure the successful implementation.

Results

Q1-Q4: Multiple Choice

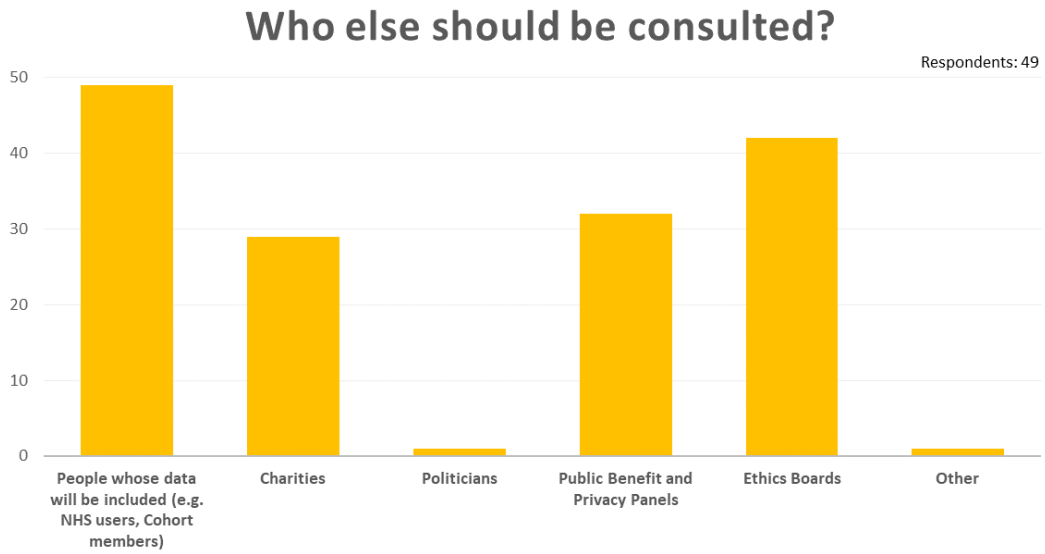
1) Who would have an interest in using this platform? *(multiple answers permitted)*



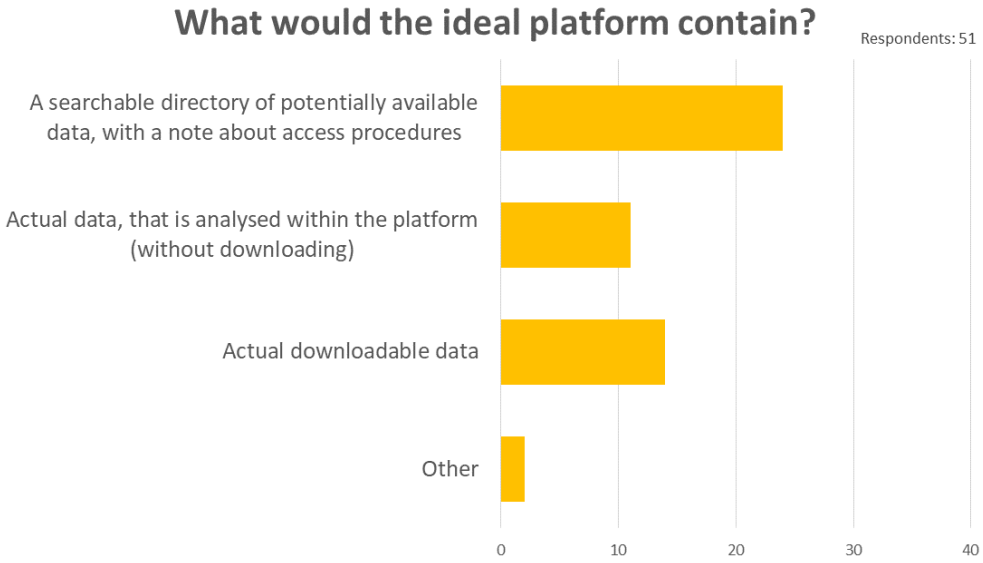
Additional Responses:

- Public Health bodies nationally and internationally
- The Public
- Patients, carers and their charities
- A range of other private companies

2) ***In addition to those listed in the last question: Who else should be consulted during its creation?*** *(multiple answers permitted)*

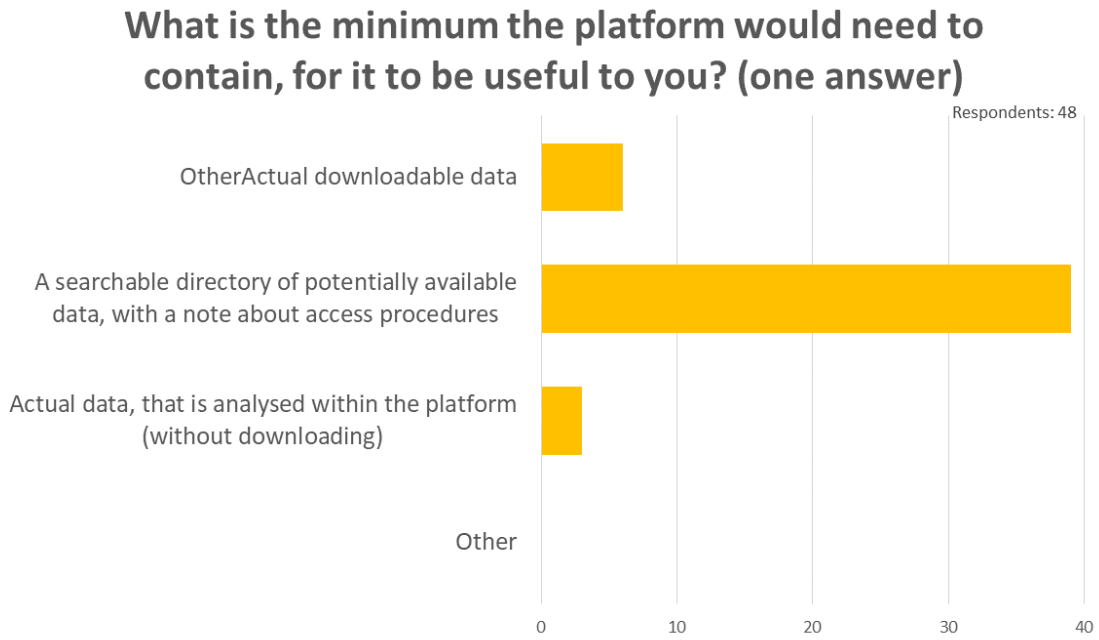


3) ***What would the ideal platform contain?*** *(one answer)*

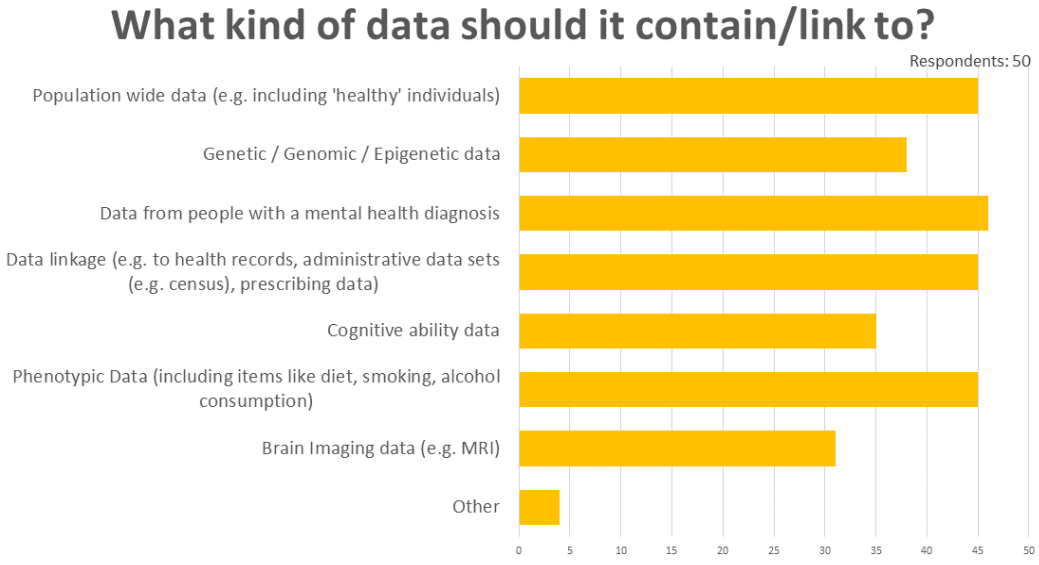


Additional Response: How can it be linked to other datasets?

4) ***What is the minimum the platform would need to contain, for it to be useful to you?*** *(one answer)*



5) ***What kind of data should it contain/link to?*** *(multiple answers permitted)*



Additional Responses:

- Prescribing data and wider social determinants of health
- Environment, social data, adversity

Q6-Q7: Free Text

Note:

1. *The examples are not direct quotes. They were drawn from the free text answers and reconsolidated to avoid repetition and improve readability.*
2. *The categories are not paralleled, and some may have crossover coverage.*

6) What research would an 'ideal' platform allow you to conduct, which is not possible now?

Category	Examples
Maximizing data usage	- Improved data linkage (e.g. ability to describe full clinical pathway) - Increased breadth of variables - Availability for re-use by others - Improved meta-data - Collation of research outputs
Broadening data capacity & accessibility	- Single site for multi-data sources - Allowing for interoperability - Augmentation of current data sets
Expanding research topics	- Coverage of minority or vulnerable groups (e.g. research around social deprivation and disparities) - Prediction analysis for scenario planning (e.g. effect of Brexit on prescription drug availability) - Ability to do research across datasets (e.g. genetic, imaging) and across disorders - If there is access to consolidated research outputs, better opportunities for collaboration between research groups and expansion of peer work - Ability to detect rare outcomes (nationwide coverage) - Incidence cohorts with follow up at scale
Conducting more rigorous research	- Using synthetic data to make an analysis plan before research design/ application for funding - Hypothesis Testing (e.g. how would this population manage without treatment, how would they respond to various interventions). - Reassurance that clinically relevant research is conducted with the right cohorts and therefore, is representative of the population - Due to improved data linkage, ability to better understand clinical pathways and how/where relevance of data lies
Improving data quality	- Reduction of bias in relation to gaps in data - Allowing for nationwide coverage (e.g. clinical and census data) - Harmonization of codes and outcome measures, making the available info more searchable - Primary to secondary care – where have patients interacted with health system, and outputs of that. - How many people actually accessing CAMS? - Secondary care prescribing – hard to access/can't find it. Collected by commercial companies – filling the gaps?

How can we maximize the platforms positive impact on people affected by mental illness?

Category	Examples
Impact/change to relevant policies	- By sharing outcomes of research based on data available from platform (e.g. to specialists, policy makers and the public) - By including James Lind Alliance Priority Setting Partnership results
Engage with patients & carers	- Increase of public awareness about available data, relevant research, intervention options, and how their data is used - Integration of engagement with patients and relevant stakeholders all throughout the research - By working with patients to increase engagement (e.g. loading up own graph and data)
Make it available for open science & public	- Obligation for open access publishing - Obligation to share methodologies - Registering protocols – eg for observational data, hub for learning, like clinical trials registry - Possibility for access for researchers worldwide - Links to biographies / communities of interest so the public can connect into research and build confidence in participating - Allow public access so the public can see descriptive results from each research output - Ensuring transparency in data we use
Improve research quality	- Establishment of feedback loops with NHS data coders - By implementing science through building on new insights derived from research done with this data - Possibility to contact researchers for clarification of findings
Tackle sociological issues	- Help fight stigma of mental illness

7) What other questions do the MRC need to ask during the scoping / consultation phase?

Category	Examples
Sustainability	- How to develop a specialized workforce for this platform? Further to that question, how can the MRC support career development for analysts in order to minimize disruption? (e.g. fixed term contracts will not be sustainable) - How to learn from best practices from other nations? (e.g. Iceland, Norway) - How to project and analyze for the future? Further to that question what will the platform look like in 20 years? - How to integrate user feedback in order to improve the platform? - How to plan for when technical infrastructure becomes obsolete? How can it be extended, how can it be gracefully retired and can some parts be used in other platforms / projects?
Qualification & training	- For the people using the platform - For the people inputting the data - Linking to PhD training via Doctoral Training Partnerships e.g. Horizon, ATI
Data management	- How to harmonize data across locations? - If / how to brainstorm with other (commercial) organizations (e.g. Tesco club card, razor database)? How to potentially link this data? - How to amplify relevance for existing datasets to be shared by researchers and participants? - How to manage research reproducibility with data from the platform? - How to make data that is donated to the platform citable?
Prioritization	- How to prioritize the different project brief elements? - Will the MRC try to encompass all possibilities or try to find a specific focus? - What can be achieved within a given time frame? - Should it be considered to start small (with a good “spine” of phenotype data) vs huge dataset with potentially large discrepancies in quality? (e.g. go small with higher levels of limited success vs going big and falling short)
Platform operation	- Requirement to think of branding on both mental health and data platform sides - Design of incentive scheme for data collectors
Governance & security	- Who would be responsible for this platform? - Who would benefit from the platform? - Who would get access? (e.g. which organizations) - Approvals process for access – single point of approval would be desirable e.g. institutional - How will access be determined e.g. cost

	- Clear, straightforward, timely process for data access. - What about considerations on standards of security? (also on international level) - How to deal with different governance standards for different data? - How to give people access to and control over their data? (e.g. learnings from care data) - Could a cross-council initiative help maximize limited resources in mental health? - How to ensure the governance policy of data access fore-sight enough?
Ethical considerations	- How to organize the process of getting consent (improve public perceptions) - How to integrate moral and ethical thinking into platform access requirements. - Data availability & purpose planning - Should there be incentives for researchers who collected the data? - How do critics of big data anticipate the potential misuse. (e.g. concerns of parents) - Consideration of planning for opt in / opt out option for children reaching adulthood - Consideration of permission and consent for mental health data collection and how it can impact research - How to address public fears on data use including re-identification?
Raise public awareness & involvement	- How can the platform be used to address mental health mis-reports on media? - How can it be used to increase the level of public understanding of genetics and their impact? - How can it be used to encourage the public to contribute? - Close engagement with patient/service user groups - Make decisions together with the communities - Under-represented participants – research needs here – may be different requirements.

Public Engagement Benefit

A national platform such as this could reduce the need for repeated public consultations.

As it stands, individual groups of researchers carry out their own engagement activities with reference to their own specific research interests.

A national platform such as this, could undertake large scale public consultations, to elucidate public opinion on research. These are likely to be more accurate (i.e. larger and more representative) and more cost effective in terms of time, effort and resources. The results could then be shared within the platform for use by multiple researchers (e.g. on their ethics and public benefit forms).

Conclusion

Our participants showed great enthusiasm for a national platform for Mental Health Data Science and could see many ways in which it would be beneficial to both researchers and a wide range of other stakeholders. However, they did express some practical and ethical concerns and agreed that a wider consultation was necessary.

Supplementary Materials:

Methodology

The MQ meeting participants were mental health data science researchers from various universities and organisations across the UK.

During the discussion session, the participants were asked to form groups (referred to as “respondents” below) in order to answer the survey collectively and anonymously. The facilitators suggested that the participants move away from their immediate colleagues or friends to avoid bias.

Groups received the survey questions as both paper handouts and on the main screen. Responses were collected through a dedicated webpage on Aha Slides, accessed via specific browser link or QR code.

Questions 1 - 5 were multiple choice questions, with questions 3 and 4 permitting only one answer. Questions 6-8 were open-ended questions.

The results were collected and analysed by the knowledge exchange team in Mental Health Data Science Scotland, under the guidance of Dr Iona Beange.

The graphs were created on excel, while the qualitative input was categorized into central themes to facilitate reading.

List of Questions

A snapshot of the questions, responses and respondent figures can be found in the table below.
(NB: one 'response' could indicate the joint discussion of several participants)

Type of question	Question	Answers	Total respondents	Total responses	Multiple answers permitted
Multiple Choice	*1) Who would have an interest in using it?*	- Researchers - Funders - Medical / pharma companies - Computing / informatics / programming companies - Clinicians - Health baords - NHS trusts - Governments - Other (collected with post-its)	46	255	yes
Multiple Choice	*2) Who else should be consulted? (in addition to those listed in the last question)*	- People whose data will be included (e.g. NHS users, cohort members) - Charities - Politicians - Public benefit & privacy panels - Ethics boards - Other (collected with post-its)	49	154	yes
Multiple Choice	*3) What would the ideal platform contain?*	- Actual downloadable data - Actual data, that is analysed within the platform (without downloading) - A searchable directory of potentially available data, with a note about access procedures - Other (collected with post-its)	51	51	no
Multiple Choice	*4) What is the minimum the platform would need to contain, for it to be useful to you?*		48	48	no
Multiple Choice	*5) What kind of data should it contain/link to?*	- Genetic / genomic / epigenetic data - Brain imaging data (e.g. MRI) - Phenotypic data (incl. Diet, smoking, alcohol consumption) - Cognitive ability data - Data linkage (e.g. to health records, admin data sets like census, prescription data) - Data from people with a mental health diagnosis - Population-wide data (e.g. incl. "Healthy" individuals) - Other (collected with post-its)	50	289	yes
Open Ended	*6) What research would an 'ideal' platform allow you to conduct, which is not possible now?*	n/a	12	24	yes

Open Ended	*7) How can we maximize the platforms positive impact on people affected by mental illness?*		15	31	yes
Open Ended	*8) What other questions do the MRC need to ask during the scoping / consultation phase?*		12	36	yes

[END report]